Published in final edited form as:
Cell. 2015 August 13; 162(4): 738–750. doi:10.1016/j.cell.2015.07.020.

Clinical sequencing uncovers origins and evolution of Lassa virus

A full list of authors and affiliations appears at the end of the article.

Summary

The 2013-2015 West African epidemic of Ebola virus disease (EVD) reminds us how little is known about biosafety level-4 viruses. Like Ebola virus, Lassa virus (LASV) can cause hemorrhagic fever with high case fatality rates. We generated a genomic catalog of almost 200 LASV sequences from clinical and rodent reservoir samples. We show that whereas the 2013-2015 EVD epidemic is fueled by human-to-human transmissions, LASV infections mainly result from reservoir-to-human infections. We elucidated the spread of LASV across West Africa and show that this migration was accompanied by changes in LASV genome abundance, fatality rates, codon adaptation, and translational efficiency. By investigating intrahost evolution, we found that mutations accumulate in epitopes of viral surface proteins, suggesting selection for immune escape. This catalog will serve as a foundation for the development of vaccines and diagnostics.

Introduction

Viruses that cause human hemorrhagic fevers, such as Ebola, Marburg and Lassa, are classified as BL-4 agents due to their high fatality rates and lack of effective treatment (Paessler and Walker, 2013). With increasing globalization, changing climatic conditions and an ever-expanding human population, our interactions with these pathogens are likely to increase (Gire et al., 2012; Lipkin, 2013). The 2013-2015 EVD epidemic (Baize et al., 2014) is a stark reminder that better understanding of these viruses is required to develop effective therapeutics and vaccines, as standard containment and isolation can be insufficient to prevent large-scale outbreaks (Pandey et al., 2014).

Corresponding author: andersen@scripps.edu (K.G.A.); happic@run.edu.ng (C.T.H.); psabeti@oeb.harvard.edu (P.C.S.).
[20] www.vhfc.org
[*] These authors contributed equally to this work
[†] These authors jointly supervised this work
[°] Deceased

Author Contributions All authors contributed in the development of this work. KGA, EP, OAF, SHK, CTH, RFG and PCS initiated the study; KGA, BJS, and PCS conceptualized the study; KGA, CBM, AEL, SKG, EME, SW, ABG, AL, CM, AI, BT-V, EMR, IO, PEE, AG, ZB, LMB, MR, and MH performed experimental work; KGA, BJS, RS, AEL, HKF, and MN analyzed the data; KGA, SKG, LMM, MS, RT, EP, LMB, SJ, DL, DSG, AG, RF, KK, MM, MF, GOA, DAA, POO, WO, IO, PEE, OF, CTH, RFG and PCS performed field work; LMM led the rodent collections; KGA, BJS, CBM, RS, SFS and PCS wrote the paper.

Lassa virus (LASV) is unique among BL-4 agents in being a common human pathogen, causing endemic disease in much of West Africa – primarily Sierra Leone, Guinea, Liberia, and Nigeria (Figure 1A). Infection with LASV can lead to acute Lassa fever (LF) with symptoms similar to EVD. LASV is estimated to hospitalize tens of thousands and cause several thousand deaths each year. Case-fatality rates (CFRs) among hospitalized LF patients can exceed 50%, although numerous sub-clinical infections are believed to occur (Troup et al., 1970; McCormick and Fisher-Hoch, 2002). Most patients are infected by exposure to excreta from the rodent *Mastomys natalensis*, which functions as a reservoir and maintains persistent infections (Lecompte et al., 2006); human-to-human transmissions have also been reported, however, primarily in hospital settings (McCormick and Fisher-Hoch, 2002; Lo Iacono et al., 2015).

LASV is a single-stranded RNA virus in the family *Arenaviridae*. Its genome consists of two segments, L (7.3kb) and S (3.4 kb), that encode four proteins: Z (matrix), L (polymerase), NP (nucleoprotein), and GPC, which is post-translationally cleaved into two peptides, GP1 and GP2, that form the transmembrane glycoprotein (Figure 1B). EBOV (*Zaire ebolavirus*) is a single-stranded RNA virus in the family *Filoviridae* with a 19 kilobase (kb) genome encoding seven proteins. While the prevalence of LASV makes it a rare model for studying the evolution of a BL-4 pathogen, only twelve whole-genome LASV sequences were available prior to this study (Djavani et al., 1997; Vieth et al., 2004).

Results

Generation of a large dataset of Lassa virus genomes

We established partnerships with Kenema Government Hospital (KGH), Sierra Leone and Irrua Specialist Teaching Hospital (ISTH), Nigeria and collected samples from LF patients between 2008 and 2013. We implemented diagnostics, training and infrastructure to ensure high quality and safe sample collection from patients hospitalized with LF (Shaffer et al., 2014).

We sequenced 183 LASV genomes from these clinical samples, eleven LASV genomes from *M. natalensis* field samples, and two genomes from viral laboratory isolates (Figure 1C and Table S1); we deposited all sequence data at NCBI (BioProject PRJNA254017) before publication. Most samples contained > 50% human material and yielded < 1% LASV reads (Figure S1A, B and Table S1). Genome coverage was fairly uniform, with higher coverage of the S than the L segment (Figure 1D), consistent with a greater copy number of S (Southern, 1996).

Since we used an unbiased sequencing approach, we were also able to assemble 7,028 unique open reading frames from the transcriptome of *M. natalensis*, a species not previously sequenced (Figure S1C-E and File S1).

Lassa virus strains are genetically diverse and cluster based on geographic location

We first examined patterns of variation and phylogenetic relationships. We found high levels of LASV nucleotide diversity, with strain variation up to 32% and 25% for the L and S segments (Figures 2A and S1F, G). This is substantially higher than previous findings based on LASV fragments (Bowen et al., 2000), and much higher than EBOV, which is more than 97% conserved across all sequenced strains (Figures 2B and S1H). We confirmed previous findings (Bowen et al., 2000) that LASV clusters into four major clades: three in Nigeria and one from the Mano River Union countries (MRU) of Sierra Leone, Guinea and Liberia (Figure 2A and Files S1-S3). We found no evidence for host-specific clades of LASV lineages; rather, samples from humans and *M. natalensis* clustered together (Figure 2A and Files S1-S3). We did not identify any recombination events within segments, but did find evidence for reassortment between segments in three samples (Figure S2A-G). This could be explained by infections of individual hosts with multiple LASV lineages followed by shuffling of segments, a process previously observed *in vitro* with LASV (Lukashevich, 1992) and *in vivo* with other arenaviruses (Stenglein et al., 2015).

Lassa virus infections are the result of multiple independent reservoir-to-human transmissions

Recent studies suggest that the 2013-2015 EVD epidemic is maintained by sustained human-to-human transmission (Gire et al., 2014) after an initial ‘spill-over’ event from a likely animal reservoir (Baize et al., 2014). Similarly, it has been suggested that up to 20% of LF cases also arise from human-to-human transmissions (Lo Iacono et al., 2015). Sustained human-to-human transmission should result in a ‘ladder-like’ structure of the phylogenetic tree along with a strong correlation between a sample's collection date and its genetic distance from the root of the tree over a short time period. Based on data from the 2013-2015 EVD epidemic (Team, 2014), we defined that time period as one year. While collection date is strongly correlated with root-to-tip distance for EBOV from the 2013-2015 EVD epidemic ($R^2 = 0.64$; Figure 2C and Table S2), the same correlation is absent for LASV sampled over a similar time period ($R^2 = 0.0001$; Figure 2D and Table S2).

Human-to-human transmission should also result in clustering of contemporaneous viral sequences on the tree. While this is pervasive across the 2013-2015 EVD epidemic samples (Gire et al., 2014) (File S1), we found that only 5 out of 169 (3%) LASV sequences from patients resulted in such clusters (Files S1 and S3). As *M. natalensis* serves as the reservoir host for LASV – and presumably maintain LASV diversity via sustained rodent-to-rodent transmission chains – we would expect rodent samples to group into more defined clusters. Indeed, 5 out of 10 (50%) LASV sequences from *M. natalensis* formed clusters consistent with rodent-to-rodent transmissions (Files S1 and S3). Finally, we also found that the average pairwise divergence for EBOV lineages in Sierra Leone from the 2013-2015 EVD epidemic was much lower than that observed for LASV lineages within individual years from Sierra Leone (Figure 2E), despite similar observed substitution rates (Figure 2F). These three lines of evidence suggest that, while EBOV during the 2013-2015 EVD epidemic is transmitted through human-to-human contact, most human LASV infections represent independent transmissions from a genetically diverse reservoir.

Lassa virus has ancient origins in modern-day Nigeria and recently spread across West Africa

While EBOV and LASV were both discovered in the latter part of the 20th century – 1976 and 1969 respectively – their origins likely vary greatly (Commission, 1978; Frame et al., 1970). Reports suggest that all EVD outbreaks share a common ancestor within the last fifty years (Carroll et al., 2013; Dudas and Rambaut, 2014; Calvignac-Spencer et al., 2014; Gire et al., 2014). In contrast, the widespread persistence of LASV in *M. natalensis,* and evidence in the human genome of natural selection linked to LASV resistance (Andersen et al., 2012), suggest that LASV might be a long-standing human pathogen.

Using molecular dating, we found that extant LASV strains likely originated in modern-day Nigeria more than a thousand years ago and spread into neighboring West African countries within the last several hundred years (Figure 2G, H). We first examined evidence for a molecular clock by comparing sample collection dates and root-to-tip distances across the entire LASV tree. In contrast to the shorter timescales analyzed above (Figure 2D), here we found significant evidence for a molecular clock ($R^2 = 0.38$, P-value < 0.0001). This allowed us to calculate the time to the most recent common ancestor (tMRCA) using Bayesian coalescent analysis (Drummond et al., 2012). We estimated the tMRCA of sampled extant LASV strains to be a little over one thousand years for the L segment (Figure S2H and Table S2; median = 1,057 years ago (ya); 915 ya - 1,218 ya, 95% Highest Posterior Density (HPD)) and 650 years for the S segment (Figure S2I and Table S2; median = 631 ya; 519 ya - 748 ya, 95% HPD). While LASV strains in Nigeria have the same tMRCA as all extant strains, those in Sierra Leone have an estimated tMRCA of only 150 years (Figure 2G, H; median = 153 ya; 137 ya - 171 ya, 95% HPD).

We tested the sensitivity of our results to key analysis parameters that could severely affect our tMRCA estimates (Wertheim and Kosakovsky Pond, 2011; Wertheim et al., 2013). We found that our estimates were robust to the choice of all tested parameters, including evolutionary model, geographical separations, and exclusion or inclusion of older ‘anchoring’ sequences, *e.g.* the 1969 Pinneo strain (Figures S3 and S4, Table S2). In linear regression of root-to-tip distance of samples on the date of collection, the sequences from the MRU showed the strongest evidence of temporal structure, suggesting that the dating is most reliable within that region (Figure S4).

Non-Nigerian Lassa virus strains have higher codon-adaptation to mammalian hosts

Previous studies have shown that viruses can adapt their codon usage to that of their hosts for translational efficiency (Sharp and Li, 1987; Bahir et al., 2009; Butt et al., 2014; Hershberg and Petrov, 2008). We examined the codon adaptation index (CAI) of LASV and EBOV to different hosts. CAI quantifies how well synonymous codon choice in the viral genome matches that of a potential host genome.

We found that LASV had a higher mean CAI than EBOV, and a similar CAI distribution across different potential mammalian hosts (Figure 3A). There was a strong linear correlation between the CAI of LASV to human and to *M. natalensis*, irrespective of which organism LASV was sequenced from (Figure S5A). In agreement with previous studies (Bahir et al., 2009), this suggests that codon adaptation to one mammal also leads to adaptation to another.

We also compared LASV sequences from patients in Sierra Leone to those from Nigeria, and found that the former had significantly higher CAI (P-value < 0.001, permutation test)

(Figure 3B). This apparent ‘burst’ of codon adaptation as LASV spread into Sierra Leone began on the branch leading out of Nigeria and remained high in most non-Nigerian strains (Figures 3C, D and S5B-E), with an even distribution across the LASV genome (Figure S5F).

As it has been suggested that dinucleotide usage play a role in determining translational efficiency of RNA viruses (Tulloch et al., 2014) we investigated whether there was a difference between Nigerian and Sierra Leonean strains, but did not observe any significant skew (Figure S5G, H).

Lassa virus genome abundance and case-fatality rates differ between Nigeria and Sierra Leone

Increased codon optimization might lead to increased viral output (Plotkin and Kudla, 2011) and therefore higher viral titers (Lauring et al., 2012) for non-Nigerian strains. With standardized inclusion criteria at our field sites (Extended Experimental Procedures), we tested this hypothesis by using qPCR to quantify LASV genome abundance. We found significantly more LASV genomes in patients from Sierra Leone than in those from Nigeria (Figure 4A). LASV genome abundance in Sierra Leone was similar to that observed in EBOV patients from the same hospital (KGH; Figure 4A) and decreased over the course of the infection (Figure S5I), likely due to treatment with the antiviral drug ribavirin (McCormick et al., 1986). Next, we binned the LASV samples into those in the top or bottom 50% CAI from within each country, and compared LASV genome abundance between bins. In Sierra Leone, individual LASV sequences with high CAI tended to have higher genome copy numbers (*P*-value < 0.05, Mann-Whitney test) but no trend was visible in Nigeria (Figure 4B, C). This suggests that CAI may affect LASV replication rate and abundance.

Since increased viremia of LASV in LF patients is correlated with higher fatality rates (McCormick and Fisher-Hoch, 2002), we might also expect CFRs to be higher in patients from Sierra Leone than from Nigeria. Again using strict criteria for inclusion, we found a significantly higher CFR (*P*-value = 0.01, Fisher's exact test) in Sierra Leonean patients than in their Nigerian counterparts (81% vs. 60%; Figure 4D). While the treatment options for LF patients are similar in the two countries, other factors could also affect genome abundances and CFRs. In particular, delay in clinical care could bias our estimates; however, self-reported times from onset of symptoms to hospital admission are the same in the two countries (average = 9.3 days; Figure 4E).

Nigerian Lassa virus strains have higher protein output than Sierra Leonean strains

Although we observed a correlation between CAI, viral genome abundance, and CFR, it remained unclear whether this is driven by differences in protein translation efficiency between Nigerian and Sierra Leonean LASV strains. We designed an experimental system to estimate translational activity for a single LASV gene with different CAI values. We randomly selected twenty LASV sequences from Nigeria and Sierra Leone, and fused the first 699 bp of their NP genes ($NP_{1\text{-}699}$) to luciferase, before cloning into expression vectors for transfection or *in vitro* translation experiments (Figure 4F). Readout of luciferase activity allowed us to detect differences in translational activity of the chimeric transcripts. As controls, we codon-optimized one LASV sequence from Nigeria and one from Sierra Leone, for an upper bound on $NP_{1\text{-}699}$-luciferase translational efficiency.

For both transfection and *in vitro* translation experiments, we observed a significant difference in translational output of the tested $NP_{1\text{-}699}$-luciferase genes, with Nigerian versions having higher outputs (Figure 4G, H). This was the opposite of the expectation based on CAI because the Sierra Leonean sequences had higher CAI (Table S2). Nigerian versions also had higher outputs for the codon-optimized forms of $NP_{1\text{-}699}$ (Figure 4H), suggesting that Nigerian sequences are intrinsically more efficient or stable.

To test whether these observations were specific to NP, we repeated the *in vitro* translation experiment using the first 736 bp of ten LASV GPC genes (Figure 4F). Once again, we found that Nigerian genes had significantly higher translational output (Figure 4I). These results suggest that there is a difference in the translational output between LASV strains from Nigeria and Sierra Leone that is independent of the variation in CAI.

Lassa virus is more diverse intrahost than Ebola virus

The long-term evolution of viruses ultimately depends on mutation and selection within individual hosts (Parameswaran et al., 2012). Our deep sequencing allowed us to examine LASV intrahost single nucleotide variants (iSNVs) within individual human and rodent hosts (Figure 5A). We called variants at a minimum minor allele frequency (minMAF) of 5% and applied stringent filtering (Extended Experimental Procedures). We validated subsets of iSNVs using different sequencing technologies and found that our results were consistent across platforms, experimental replicates, library preparations, and variant calling methods (Figures S5J-L and S6).

We found that *M. natalensis* generally harbor more LASV iSNVs than humans (median iSNVs/kb = 1.5 vs. 0.1; *P*-value < 0.0001, Mann-Whitney test), consistent with longer, more chronic infections (Figures 5B and S7A-D). LASV is significantly more diverse intrahost than EBOV (accounting for differences in sequence coverage between the two – median bp coverage ~ 2,000× for EBOV (Gire et al., 2014) and ~ 250× for LASV, Figure 1C; *P*-value = 0.0005, Mann-Whitney test) with an average number of iSNVs per covered site of 2.1×10^{-3} in LF patients, but only 1.3×10^{-4} in EVD patients (Figure 5C). This difference is primarily driven by a subset of LASV-infected individuals that have >15 iSNVs – diversity similar to that observed in *M. natalensis* (Figure 5C). Such high diversity – with iSNV frequencies that appear stable over the course of infection (Figure S7E) – was never observed in EVD patients (Figure 5B, C).

LF is generally considered an acute disease in humans (McCormick and Fisher-Hoch, 2002), but high numbers of iSNVs could be explained by long-term chronic infections and/or adaptive evolution of LASV. An alternative explanation is multiple infections; however, the wide range of allele frequencies (Figure S7F) and general lack of linkage between iSNVs (Table S2) argues against this being the prevailing explanation. In addition, the vast majority of iSNVs (94.4%) are transitions, rather than transversion mutations (Figure S5L), which accumulate over longer evolutionary timescales (Wakeley, 1996). This suggests that most iSNV are evolutionarily recent, and that LASV iSNVs arise mostly via *de novo* mutation within hosts, and more rarely via transmission and multiple infections by circulating strains.

Natural selection is acting on the Lassa virus glycoprotein

We next investigated the role of natural selection in shaping intrahost variation. In LASV, we observed a significantly higher dN/dS (*P*-value = 0.0013, Permuted McDonald-Kreitman test), a measure of selective constraint at the protein level, within hosts than between hosts (Figure 5D). For EBOV, the trend was in the same direction, but not statistically significant (Figure 5D). Assuming that dN represents mostly deleterious mutations (Shapiro et al., 2009), this is consistent with these mutations being purged by purifying selection over evolutionary time (Rocha et al., 2006). Because purifying selection has less time to act within a single host, dN/dS is higher within hosts than between. However, the dN/dS of ~0.2 for LASV iSNVs is still much less than ~1 expected in the absence of any selection (Anisimova and Liberles, 2007). In contrast, iSNVs in certain other viruses such as dengue approach the neutral expectation of dN/dS ~ 1 (Holmes, 2003). Also reflecting purifying selection on LASV within hosts, the dN/dS ratio appears to decrease at higher iSNV frequencies (Figure S7G). EBOV intrahost dN/dS is higher than LASV (Figure 5D), consistent with LASV intrahost populations being subject to stronger (or a longer duration of) purifying selection.

Intrahost dN/dS varied widely across LASV genes, suggesting different selective pressures on individual genes. Most notably, GPC genes sequenced from both human and *M. natalensis*, had a significantly higher dN/dS ratio within hosts than between hosts (Figure 5E). GPC encodes the only protein partially exposed on the outside of the LASV particle (Figure 1B). It has a significantly higher within-host dN/dS than NP (*P*-value < 0.05, Fisher's exact test), the neighboring gene on the S segment (Figure 5E), but similar between-host dN/dS. These results suggest either a GPC-specific relaxation of within-host purifying selection or within-host diversifying (positive) selection (Baum et al., 2003).

Nonsynonymous Lassa virus intrahost variants accumulate in predicted epitopes in the glycoprotein

We hypothesized that immune pressures on LASV GPC could drive within-host diversifying selection, favoring nonsynonymous iSNVs that impair immune detection by disrupting epitopes. This phenomenon has been reported for other viruses; *e.g.* at the population-wide level in pandemic influenza A virus (Bhatt et al., 2011). To evaluate whether iSNVs disrupt epitopes, we used a machine learning method (El-Manzalawy et al., 2008) to predict linear B cell epitopes in each LASV protein.

Nonsynonymous iSNVs in LASV GPC occurred in predicted B cell epitopes significantly more than expected by chance (Figure 6A, B; *P*-value < 0.01, Binomial test). This was true for LASV samples from patients and *M. natalensis* independently, although the signal was stronger in patients (Figure 6A). In contrast, synonymous iSNVs were randomly distributed across GPC, consistent with their lack of impact on epitope structure (Figure 6A, B). We observed a similar but weaker trend for NP, although this difference only reached statistical significance in *M. natalensis* (Figure 6A).

To test if nonsynonymous iSNVs interfere with B cell epitope recognition, we reran the B cell epitope predictions, changing single amino acids within the epitopes from the consensus call to the iSNV variant. For 14 of the 18 predicted B cell epitopes, changing the iSNV from the consensus to the variant allele significantly reduced the epitope score (Table S2; *P*-value = 0.015, Sign test).

To test if nonsynonymous iSNVs also appear to fall within T cell epitopes, we predicted T cell epitopes in each LASV protein (Extended Experimental Procedures). We found that nonsynonymous iSNVs accumulated to some extent in LASV GPC, although the results did not reach statistical significance (Figure 6C; *P*-value = 0.07, Binomial test).

Intrahost variants interfere with antibody binding

To investigate the functional effects of a subset of LASV iSNVs, we created iSNV mutations in predicted B cell epitopes in GP1 (Extended Experimental Procedures), expressed them in HEK293 cells, and tested their binding to a panel of GPC-specific monoclonal antibodies (mAbs) using flow cytometry. These mutations led to a significant drop in the average mean fluorescence intensity (MFI) for GP1-specific mAbs (Figure 6D), consistent with diminished mAb binding. Similarly, when we investigated the effects of single point-mutations within GP1 epitopes, we found that minor alleles in the LASV population displayed significantly reduced binding to GP1-specific mAbs (Figure 6E).

These observations suggest that the host adaptive immune system imposes selective pressures on the intrahost viral population, driving an accumulation of nonsynonymous iSNVs in LASV GPC.

Nonsynonymous Lassa virus intrahost variants tend not to become fixed in other hosts

To further explore the evolution of LASV within and between hosts, we investigated how often iSNVs become fixed in other consensus sequences. We defined an iSNV as ‘fixed’ if its minor allelic variant was observed in one or more LASV consensus sequences. We observed a significantly higher nonsynonymous to synonymous ratio (N/S) for unfixed compared to fixed iSNVs (Figure 7A), suggesting a selective bias against the fixation of nonsynonymous iSNVs. LASV and EBOV both have similar numbers of unfixed iSNVs, but LASV has many more fixed iSNVs, likely due to higher rates of iSNV fixation (or transmission) in LASV than EBOV (Figure 7B). However, the putative transmitted (fixed) iSNVs tend to be biased toward synonymous mutations. This bias is much stronger in LASV (Figure 7C, top panel) but still detectable in EBOV (Figure 7C, middle panel). The bias cannot be attributed to differences in minor allele frequencies between nonsynonymous and synonymous iSNVs (*P*-value > 0.1, Kolmogorov-Smirnov test), or to a correlation between MAF and prevalence in consensus sequences (*P*-value > 0.1 for both N and S, Pearson's correlation); therefore, it is best attributed to selection against transmission and/or fixation of nonsynonymous iSNVs.

A single suspected transmission event, between a pair of *M. natalensis* captured on the same day from the same household, provided an opportunity to observe iSNV fixation dynamics on short timescales. The two samples, Z0947 and Z0948, are nearest-neighbors on the

LASV phylogeny (Files S1-S3), suggesting recent (but not necessarily direct) transmission (Extended Experimental Procedures). Assuming that transmission occurred from Z0948 to Z0947, we observed that derived alleles reaching high frequency (DAF > 0.5) in Z0947 tended to be nonsynonymous, while derived alleles remaining at lower frequency (DAF < 0.5) were always synonymous (Figure 7C, bottom panel). Other transmission scenarios (Figure S7H and Extended Experimental Procedures) also confirm that nonsynonymous iSNVs reach high frequency within a host, but fail to be transmitted to the next host. Along with the dN/dS and epitope analyses, this supports a model in which nonsynonymous iSNVs rise to high frequency within an individual due to positive selection, but are less likely to become fixed in other hosts due to purifying selection.

Discussion

Comparing Ebola virus and Lassa virus evolutionary dynamics

EBOV and LASV are RNA viruses that can lead to illnesses with similar clinical symptoms, yet they differ markedly in their epidemiology and evolutionary dynamics. LASV is more than an order of magnitude more diverse than EBOV (Figure 2B), and molecular dating suggests that it has been circulating in Nigeria for over a thousand years, followed by a more recent spread across West Africa (Figure 2G). In contrast, it has been suggested that the Makona variant of EBOV responsible for the EVD epidemic in West Africa was introduced over the last decade (Dudas and Rambaut, 2014; Calvignac-Spencer et al., 2014; Gire et al., 2014).

These analyses, however, provide lower-bound tMRCA estimates of sampled (extant) viral lineages; the true ages of all LASV and EBOV lineages are likely much older (Taylor et al., 2010). Due to limited sampling from Guinea and Liberia, our LASV dating analysis is likely most accurate within Sierra Leone, although we achieved comparable results when considering the entire dataset or the individual regions alone (Figure S3E). Furthermore, our 400 year old ‘out of Nigeria’ estimate relies on a single sequence from the Ivory Coast; additional sampling outside the MRU and Nigeria could push this date back.

Due to the high heterogeneity among LASV lineages, continuous monitoring of its mutational spectrum and evolutionary change will be critical for the development of effective vaccines and diagnostics. Since LASV strains cluster by geography, it is more conserved within individual countries. For example, average sequence identity among lineages from Sierra Leone is 90% at the nucleotide level and 95% at the amino acid level (Figure S1F). A useful strategy might therefore be to develop diagnostics, vaccines, and sequence-specific therapeutics that are country-specific, or that target the most conserved features of the viral genome.

The 2013-2015 West African EVD epidemic likely originated from a single zoonotic transmission event (Baize et al., 2014), followed by sustained human-to-human transmission and clock-like, linear accumulation of mutations (Figure 2C). In contrast, LASV has a clock-like signature on the timescale of decades (Figure S4B), but not on shorter timescales (Figure 2D). Combined with the intermingling of human and *M. natalensis* samples on the phylogenetic tree (Figure 2A), this is consistent with a genetically diverse pool of LASV being maintained in its rodent reservoir, with most human infections caused by genetically distinct viruses. A recent study suggested that human-to-human transmission of LASV may account for up to 20% of all cases (Lo Iacono et al., 2015), but we found little support for this in our dataset. This does not rule out human-to-human transmission entirely, but it suggests that human transmission chains are the exception rather than the rule.

LASV is more polymorphic within hosts than EBOV, and *M. natalensis* hosts harbor more polymorphic LASV populations than humans (Figure 5B, C). Since most LF and EVD patients have 0 or 1 iSNVs, the difference difference between LASV and EBOV is mostly driven by a subset of LF patients with many LASV iSNVs (Figure 5B, C). LASV iSNV frequencies tend to remain stable over the relatively short period of hospitalization (Figure S7E), suggesting that intrahost *de novo* mutations and frequency changes may take time to develop, or may occur early in the infection. These observations suggest that – at least in some patients –LASV infections could last longer than EBOV infections, allowing more time for the generation of polymorphism within hosts.

Longer infections also provide more time for natural selection to eliminate deleterious mutations from the viral population. Consistent with longer infection periods in LASV, dN/dS ratios are lower within LF patients than EVD patients (Figure 5D).

While these findings are consistent with the existence of chronic LASV infections in humans, they do not constitute proof. Further studies are needed to verify the causes of high-diversity LASV infection and the prevalence of non-acute human infections. Compelling evidence could come from longitudinal sampling of asymptomatic carriers of LASV, for example.

Codon adaptation, translational efficiency and genome abundance

We uncovered significant differences in genome abundance, CFR, CAI, and translational efficiency of LASV strains from Sierra Leone and from Nigeria. The increase in CAI of non-Nigerian strains (Figure 3B-D), along with higher viral copy numbers and CFRs in human patients (Figure 4A-D), would seem to suggest that the virus evolved towards greater human virulence. There are indeed many examples suggesting that pathogens with natural reservoirs may evolve toward greater human virulence, so long as they remain avirulent in the reservoir host (Ewald, 1994). We have not, however, been able to establish causality among these observations. While clinical care for LF patients is similar at ISTH and KGH (Extended Experimental Procedures) – and time from symptom onset to hospitalization appears to be the same (Figure 4E) – several additional parameters are beyond our control, including variations in socioeconomic, clinical, and human genetic factors between Sierra Leone and Nigeria. These prevent us from determining whether the difference in CFRs has a viral genetic basis and whether variations in CAI, viral genome abundance, and CFRs are causally linked. In addition, while we observed significantly higher LASV genome abundance in LF patients from Sierra Leone, we could not determine whether this difference translates into higher infectious titers. Controlled animal studies in BL-4 laboratories comparing strains from the two countries would help resolve this important question.

We expected that the increased CAI of Sierra Leonean LASV lineages would lead to increased translational output in an experimental system. Surprisingly, the opposite was true. Nigerian LASV lineages had significantly higher *in vitro* protein outputs than their Sierra Leonean counterparts (Figure 4G-I). Translation-independent mechanisms – such as post-translational modifications and protein stability – could explain these observations, irrespective of CAI differences.

If the progenitor to Sierra Leone LASV strains indeed had lower translational output, the emergence of LASV strains with increased CAI could have been driven by positive selection to compensate. The higher CAI could then have led to higher viral titers in human patients. Alternatively, the increase in CAI could have been caused by genetic drift. Under this scenario, mutational biases (Extended Experimental Procedures) and drift – combined with insufficient time for mutations to be exposed to purifying selection in Sierra Leone – led to an increase in CAI independently of positive selection. The Sierra Leonean tMRCA of the LASV population is indeed relatively recent (Figure 2G), having possibly undergone a recent population bottleneck (Lalis et al., 2012). Further work will be required to disentangle the adaptive and neutral contributions to the evolution of CAI, and whether changes in CAI affect viral fitness and CFRs.

Immune selection on Lassa virus within hosts

The observation that LASV GPC has the highest intrahost dN/dS (Figure 5E) and the most nonsynonymous iSNVs in predicted epitopes (Figure 6) suggests that it is a target of immune-driven diversifying selection within hosts, in both humans and *M. natalensis*. This effect was most strongly supported when we investigated the involvement of B cells (Figure 6A), although we also found weaker evidence for the role of T cells (Figure 6C).

Given that human LASV infections are typically thought to be of short duration, it is perhaps surprising that there would be enough time for the host to mount an antibody response, and for viral diversity to be measurably shaped by this selective pressure. However, this is not entirely implausible, since it is known that LASV-specific antibodies appear relatively quickly, with experimentally infected nonhuman primates developing detectable IgM titers after nine days and IgG titers after twelve (Baize et al., 2009). Furthermore, it has also been shown that 25% of LF patients in Sierra Leone are IgM-positive upon admission and 10% are IgG-positive (Shaffer et al., 2014). Combined with the observation that several LF patients have LASV iSNV numbers similar to those observed in *M. natalensis* (Figure 5B, C), sufficient time for antibody-mediated responses has likely elapsed in a number of LF patients when they present at the hospital. We were unable to test seropositivity in our cohort, but future studies could assess whether there is a correlation between the presence or absence of LASV-specific antibodies and the number of iSNVs. Our prediction would be that LF patients with high numbers of LASV iSNVs should also tend to be IgM- and/or IgG-positive upon admission.

Conclusions

Our dataset of full-length LASV genomes yields insights into the biogeography, evolution and spread of a hemorrhagic fever-causing virus. Further improvements in collecting, sequencing, and analyzing LASV and EBOV samples combined with concurrent experiments in BL-4 laboratories will allow us to pursue open areas of inquiry. The catalogue of LASV genetic diversity presented here is thus an essential foundation for the development of vaccines and diagnostics for a highly diverse, continuously evolving, and unusually prevalent BL-4 agent.

Experimental Procedures

Ethics statement

Subjects were recruited for this study using protocols approved by relevant human subjects committees. All patients were treated with a similar standard of care.

Samples

All samples were acquired on the day of admission before any treatment regimens had begun. Ten ml of whole blood was collected and plasma or serum was prepared. Diagnostic tests for the presence of LASV were performed on-site. Rodents (all from Sierra Leone) were trapped in case-households, humanely sacrificed and samples were collected from serum and/or spleen. All samples were inactivated in either Buffer AVL (Qiagen) or TRIzol (Life Technologies) following the manufacturer's standard protocols and stored in solar-driven −20° C freezers.

Case-fatality rates

CFRs were calculated from patients that had all of the following characteristics: (i) known outcome, (ii) positive for LASV in the field, (iii) confirmed positive upon retesting in the US, and (iv) sequencing confirmed the presence of LASV.

Sequencing

cDNA synthesis and Illumina library construction were performed, libraries were pooled, and sequenced on the Illumina platform. All the data were deposited at NCBI (BioProject PRJNA254017). EBOV data is available under PRJNA257197.

Assembly of LASV genomes

LASV genomes were *de novo* assembled followed by read mapping and duplicate removal.

Trees

Maximum likelihood phylogenies were made with RAxML v7.3.0 (Stamatakis et al., 2005) and rooted using the 1969 Pinneo strain.

Molecular dating

BEAST v1.7.4 (Drummond and Rambaut, 2007) was used with a model incorporating a lognormal relaxed clock, exponential growth, HKYγi with four categories and codon partitioning (srd06).

Intra-host variants

iSNV were called and filtered taking into consideration minimum coverage, minimum frequency (5%), strand-bias, and base quality.

Intra-host selection

A modified version of the McDonald-Kreitman test (McDonald and Kreitman, 1991) was applied to assess evidence for natural selection.

Epitope prediction

B cell epitopes were predicted using BCPRED(El-Manzalawy et al., 2008). T cell epitopes were predicted using NetCTL (Larsen et al., 2007).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Kristian G. Andersen[1,2,3,*], B. Jesse Shapiro[1,2,4,*], Christian B. Matranga[2,*], Rachel Sealfon[2,5], Aaron E. Lin[1,2], Lina M. Moses[6], Onikepe A. Folarin[7,8], Augustine Goba[9], Ikponmwonsa Odia[7], Philomena E. Ehiane[7], Mambu Momoh[9,10], Eleina M. England[2], Sarah Winnicki[1,2], Luis M. Branco[11], Stephen K. Gire[1,2], Eric Phelan[2], Ridhi Tariyal[2], Ryan Tewhey[1,2], Omowunmi Omoniwa[7], Mohammed Fullah[9,10,°], Richard Fonnie[8,°], Mbalu Fonnie[9,°], Lansana Kanneh[9], Simbirie Jalloh[9], Michael Gbakie[9], Sidiki Saffa[9,°], Kandeh Karbo[9], Adrianne D. Gladden[2], James Qu[2], Matthew Stremlau[1,2], Mahan Nekoui[1,2], Hilary K. Finucane[2], Shervin Tabrizi[1,2], Joseph J. Vitti[1], Bruce Birren[2], Michael Fitzgerald[2], Caryn McCowan[2], Andrea Ireland[2], Aaron M. Berlin[2], James Bochicchio[2], Barbara Tazon-Vega[2], Niall J. Lennon[2], Elizabeth M. Ryan[2], Zach Bjornson[12], Danny A. Milner JR[13], Amanda K. Lukens[13], Nisha Broodie[14], Megan Rowland[11], Megan Heinrich[11], Marjan Akdag[11], John S. Schieffelin[6], Danielle Levy[6], Henry Akpan[15], Daniel G. Bausch[6], Kathleen Rubins[16], Joseph B. McCormick[17], Eric S. Lander[2], Stephan Günther[18], Lisa Hensley[19], Sylvanus Okogbenin[7], Viral Hemorrhagic Fever Consortium[20], Stephen F. Schaffner[2], Peter O. Okokhere[7], S. Humarr Khan[9,°], Donald S. Grant[9], George O. Akpede[7], Danny A. Asogun[7], Andreas Gnirke[2], Joshua Z. Levin[2,†], Christian T. Happi[7,8,†], Robert F. Garry[6,†], and Pardis C. Sabeti[1,2,13,†]

Affiliations

[1]FAS Center for Systems Biology, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

[2]Broad Institute, Cambridge, MA 02142, USA

[3]The Scripps Research Institute, Scripps Translational Science Institute, La Jolla, CA 92037,USA

[4]Department of Biological Sciences, University of Montréal, Montréal, QC H2V 2S9, Canada

[5]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[6]Tulane Health Sciences Center, Tulane University, New Orleans, LA 70118, USA

[7]Institute of Lassa Fever Research and Control, Irrua Specialist Teaching Hospital, Irrua, Nigeria

[8]Department of Biological Sciences, College of Natural Sciences, Redeemer's University, Redemption City, Nigeria

[9]Lassa Fever Laboratory, Kenema Government Hospital, Kenema, Sierra Leone

[10]Eastern Polytechnic College, Kenema, Sierra Leone

[11]Zalgen Labs, Germantown, MD 20876, USA

[12]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94304, USA

[13]Department of Immunology and Infectious Disease, Harvard School of Public Health, Boston, MA 02115, USA

[14]College of Medicine, Columbia University, New York, NY 10032, USA

[15]Nigerian Federal Ministry of Health, Abuja, Nigeria

[16]The National Aeronautics and Space Administration, Johnson Space Center, TX 77058, USA

[17]The University of Texas School of Public Health, Brownsville, TX 77030, USA

[18]Department of Virology, Bernhard-Nocht-Institute for Tropical Medicine, 20259 Hamburg, Germany

[19]NIAID Integrated Research Facility, Frederick, MD 21702, USA

Acknowledgments

We thank C. Edwards, X. Yang, M. Zody, B. Han, A. Andersen, I. Shlyakhter, D. Park, M. Henn, M. Busby, M. Kellis, C. Bishop, A. Haislip, L. Burchfield, A. Matthews, and F. Viloria for their help and advice. We also thank the Ministry of Health & Sanitation Government of Sierra Leone, Oyo State Ministry of Health and the Federal Ministry of Health, Nigeria. This project has been funded with Federal funds from the NIH, Department of Health and Human Services, under award 1DP2OD006514-01 and NIAID Contracts: HHSN272200900049C, HHSN272201000022C, HHSN272200900018C, and U19AI110818. We have received additional support from USAMRAA W81XWH-10-1-0098, a Packard Foundation Fellowship for Science and Engineering, a Broad Institute SPARC award and GU 883/1-1 from the German Research Foundation. KGA was supported by a fellowship from the Carlsberg Foundation, BJS by a Harvard MIDAS CCDD postdoctoral fellowship and Canada Research Chair, and by NSF GRFP fellowships for RS and AEL with Grant Nos. DGE 1122374 and 1144152.

References

Andersen KG, Shylakhter I, Tabrizi S, Grossman SR, Happi CT, Sabeti PC. Genome-wide scans provide evidence for positive selection of genes implicated in Lassa fever. Philos Trans R Soc Lond B Biol Sci. 2012; 367:868–877. [PubMed: 22312054]

Anisimova M, Liberles DA. The quest for natural selection in the age of comparative genomics. Heredity (Edinb). 2007; 99:567–579. [PubMed: 17848974]

Bahir I, Fromer M, Prat Y, Linial M. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. Mol Syst Biol. 2009; 5:311. [PubMed: 19888206]

Baize S, Marianneau P, Loth P, Reynard S, Journeaux A, Chevallier M, Tordo N, Deubel V, Contamin H. Early and strong immune responses are associated with control of viral replication and recovery in lassa virus-infected cynomolgus monkeys. J Virol. 2009; 83:5890–5903. [PubMed: 19297492]

Baize S, Pannetier D, Oestereich L, Rieger T, Koivogui L, Magassouba N, Soropogui B, Sow MS, Keïta S, De C Hilde, Tiffany A, Dominguez G, Loua M, Traoré A, Kolié M, Malano ER, Heleze E, Bocquin A, Mély S, Raoul H, Caro V, Cadar D, Gabriel M, Pahlmann M, Tappe D, Schmidt-Chanasit J, Impouma B, Diallo AK, Formenty P, Van H Michel, Günther S. Emergence of Zaire Ebola Virus Disease in Guinea —Preliminary Report. N Engl J Med. 2014 140416140039002.

Baum J, Thomas AW, Conway DJ. Evidence for diversifying selection on erythrocyte-binding antigens of Plasmodium falciparum and P. vivax. Genetics. 2003; 163:1327–1336. [PubMed: 12702678]

Bhatt S, Holmes EC, Pybus OG. The genomic rate of molecular adaptation of the human influenza A virus. Mol Biol Evol. 2011; 28:2443–2451. [PubMed: 21415025]

Bowen MD, Rollin PE, Ksiazek TG, Hustad HL, Bausch DG, Demby AH, Bajani MD, Peters CJ, Nichol ST. Genetic diversity among Lassa virus strains. J Virol. 2000; 74:6992–7004. [PubMed: 10888638]

Butt AM, Nasrullah I, Tong Y. Genome-wide analysis of codon usage and influencing factors in chikungunya viruses. PLoS One. 2014; 9:e90905. [PubMed: 24595095]

Calvignac-Spencer S, Schulze JM, Zickmann F, Renard BY. Clock Rooting Further Demonstrates that Guinea 2014 EBOV is a Member of the Zaire Lineage. PLoS Curr. 2014; 6

Carroll SA, Towner JS, Sealy TK, McMullan LK, Khristova ML, Burt FJ, Swanepoel R, Rollin PE, Nichol ST. Molecular evolution of viruses of the family Filoviridae based on 97 whole-genome sequences. J Virol. 2013; 87:2608–2616. [PubMed: 23255795]

Djavani M, Lukashevich IS, Sanchez A, Nichol ST, Salvato MS. Completion of the Lassa fever virus sequence and identification of a RING finger open reading frame at the L RNA 5′ End. Virology. 1997; 235:414–418. [PubMed: 9281522]

Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007; 7:214. [PubMed: 17996036]

Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012; 29:1969–1973. [PubMed: 22367748]

Dudas G, Rambaut A. Phylogenetic Analysis of Guinea 2014 EBOV Ebolavirus Outbreak. PLoS Curr. 2014; 6

El-Manzalawy Y, Dobbs D, Honavar V. Predicting linear B-cell epitopes using string kernels. J Mol Recognit. 2008; 21:243–255. [PubMed: 18496882]

Ewald, PW. Evolution of Infectious Disease. New York: Oxford University Press; 1994.

Fichet-Calvet E, Rogers DJ. Risk maps of Lassa fever in West Africa. PLoS Negl Trop Dis. 2009; 3:e388. [PubMed: 19255625]

Frame JD, Baldwin JMJ, Gocke DJ, Troup JM. Lassa fever, a new virus disease of man from West Africa. I. Clinical description and pathological findings. Am J Trop Med Hyg. 1970; 19:670–676. [PubMed: 4246571]

Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, Kanneh L, Jalloh S, Momoh M, Fullah M, Dudas G, Wohl S, Moses LM, Yozwiak NL, Winnicki S, Matranga CB, Malboeuf CM, Qu J, Gladden AD, Schaffner SF, Yang X, Jiang PP, Nekoui M, Colubri A, Coomber MR, Fonnie M, Moigboi A, Gbakie M, Kamara FK, Tucker V, Konuwa E, Saffa S, Sellu J, Jalloh AA, Kovoma A, Koninga J, Mustapha I, Kargbo K, Foday M, Yillah M, Kanneh F, Robert W, Massally JL, Chapman SB, Bochicchio J, Murphy C, Nusbaum C, Young S, Birren BW, Grant DS, Scheiffelin JS, Lander ES, Happi C, Gevao SM, Gnirke A, Rambaut A, Garry RF, Khan SH, Sabeti PC. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science. 2014; 345:1369–1372. [PubMed: 25214632]

Gire SK, Stremlau M, Andersen KG, Schaffner SF, Bjornson Z, Rubins K, Hensley L, McCormick JB, Lander ES, Garry RF, Happi C, Sabeti PC. Epidemiology. Emerging disease or diagnosis? Science. 2012; 338:750–752. [PubMed: 23139320]

Hershberg R, Petrov DA. Selection on codon bias. Annu Rev Genet. 2008; 42:287–299. [PubMed: 18983258]

Holmes EC. Patterns of intra- and interhost nonsynonymous variation reveal strong purifying selection in dengue virus. J Virol. 2003; 77:11296–11298. [PubMed: 14512579]

Lalis A, Leblois R, Lecompte E, Denys C, Ter Meulen J, Wirth T. The impact of human conflict on the genetics of Mastomys natalensis and Lassa virus in West Africa. PLoS One. 2012; 7:e37068. [PubMed: 22615894]

Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics. 2007; 8:424. [PubMed: 17973982]

Lauring AS, Acevedo A, Cooper SB, Andino R. Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. Cell Host Microbe. 2012; 12:623–632. [PubMed: 23159052]

Lecompte E, Fichet-Calvet E, Daffis S, Koulemou K, Sylla O, Kourouma F, Dore A, Soropogui B, Aniskin V, Allali B, Kouassi Kan S, Lalis A, Koivogui L, Gunther S, Denys C, ter Meulen J. Mastomys natalensis and Lassa fever, West Africa. Emerg Infect Dis. 2006; 12:1971–1974. [PubMed: 17326956]

Lipkin WI. The changing face of pathogen discovery and surveillance. Nat Rev Microbiol. 2013; 11:133–141. [PubMed: 23268232]

Lo Iacono G, Cunningham AA, Fichet-Calvet E, Garry RF, Grant DS, Khan SH, Leach M, Moses LM, Schieffelin JS, Shaffer JG, Webb CT, Wood JL. Using modelling to disentangle the relative contributions of zoonotic and anthroponotic transmission: the case of lassa Fever. PLoS Negl Trop Dis. 2015; 9:e3398. [PubMed: 25569707]

Lukashevich IS. Generation of reassortants between African arenaviruses. Virology. 1992; 188:600–605. [PubMed: 1585636]

McCormick JB, Fisher-Hoch SP. Lassa fever. Curr Top Microbiol Immunol. 2002; 262:75–109. [PubMed: 11987809]

McCormick JB, King IJ, Webb PA, Scribner CL, Craven RB, Johnson KM, Elliott LH, Belmont-Williams R. Lassa fever. Effective therapy with ribavirin. N Engl J Med. 1986; 314:20–26. [PubMed: 3940312]

McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 1991; 351:652–654. [PubMed: 1904993]

Paessler S, Walker DH. Pathogenesis of the viral hemorrhagic fevers. Annu Rev Pathol. 2013; 8:411–440. [PubMed: 23121052]

Pandey A, Atkins KE, Medlock J, Wenzel N, Townsend JP, Childs JE, Nyenswah TG, Ndeffo-Mbah ML, Galvani AP. Strategies for containing Ebola in West Africa. Science. 2014; 346:991–995. [PubMed: 25414312]

Parameswaran P, Charlebois P, Tellez Y, Nunez A, Ryan EM, Malboeuf CM, Levin JZ, Lennon NJ, Balmaseda A, Harris E, Henn MR. Genome-wide patterns of intrahuman dengue virus diversity reveal associations with viral phylogenetic clade and interhost diversity. J Virol. 2012; 86:8546–8558. [PubMed: 22647702]

Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet. 2011; 12:32–42. [PubMed: 21102527]

Commission, R. O. A. I. Ebola haemorrhagic fever in Zaire, 1976. Bull World Health Organ. 1978; 56:271–293. [PubMed: 307456]

Rocha EP, Smith JM, Hurst LD, Holden MT, Cooper JE, Smith NH, Feil EJ. Comparisons of dN/dS are time dependent for closely related bacterial genomes. J Theor Biol. 2006; 239:226–235. [PubMed: 16239014]

Shaffer JG, Grant DS, Schieffelin JS, Boisen ML, Goba A, Hartnett JN, Levy DC, Yenni RE, Moses LM, Fullah M, Momoh M, Fonnie M, Fonnie R, Kanneh L, Koroma VJ, Kargbo K, Ottomassathien D, Muncy IJ, Jones AB, Illick MM, Kulakosky PC, Haislip AM, Bishop CM, Elliot DH, Brown BL, Zhu H, Hastie KM, Andersen KG, Gire SK, Tabrizi S, Tariyal R, Stremlau M, Matschiner A, Sampey DB, Spence JS, Cross RW, Geisbert JB, Folarin OA, Happi CT, Pitts KR, Geske FJ, Geisbert TW, Saphire EO, Robinson JE, Wilson RB, Sabeti PC, Henderson LA, Khan SH, Bausch DG, Branco LM, Garry RF. Lassa fever in post-conflict sierra leone. PLoS Negl Trop Dis. 2014; 8:e2748. [PubMed: 24651047]

Shapiro BJ, David LA, Friedman J, Alm EJ. Looking for Darwin's footprints in the microbial world. Trends Microbiol. 2009; 17:196–204. [PubMed: 19375326]

Sharp PM, Li WH. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 1987; 15:1281–1295. [PubMed: 3547335]

Southern, P. Arenaviridae: The Viruses and Their Replication. In: Fields, BN.; Knipe, DM.; Howley, PM., editors. Fundamental Virology. Philadelphia: Lippincott-Raven; 1996. p. 675-690.

Stamatakis A, Ludwig T, Meier H. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics. 2005; 21:456–463. [PubMed: 15608047]

Stenglein MD, Jacobson ER, Chang LW, Sanders C, Hawkins MG, Guzman DS, Drazenovich T, Dunker F, Kamaka EK, Fisher D, Reavill DR, Meola LF, Levens G, DeRisi JL. Widespread recombination, reassortment, and transmission of unbalanced compound viral genotypes in natural arenavirus infections. PLoS Pathog. 2015; 11:e1004900. [PubMed: 25993603]

Taylor DJ, Leach RW, Bruenn J. Filoviruses are ancient and integrated into mammalian genomes. BMC Evol Biol. 2010; 10:193. [PubMed: 20569424]

Troup JM, White HA, Fom AL, Carey DE. An outbreak of Lassa fever on the Jos plateau, Nigeria, in January-February 1970. A preliminary report. Am J Trop Med Hyg. 1970; 19:695–696. [PubMed: 4987549]

Tulloch F, Atkinson NJ, Evans DJ, Ryan MD, Simmonds P. RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. Elife. 2014; 4

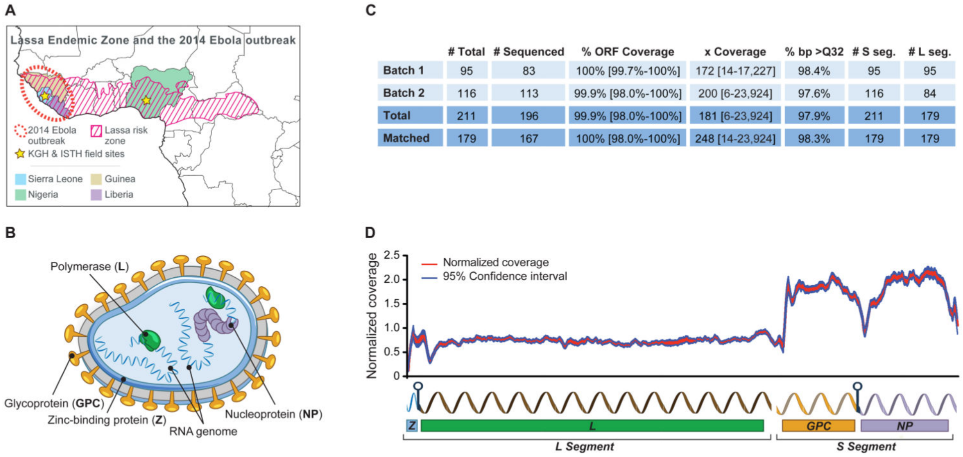
Vieth S, Torda AE, Asper M, Schmitz H, Gunther S. Sequence analysis of L RNA of Lassa virus. Virology. 2004; 318:153–168. [PubMed: 14972544]

Wakeley J. The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. Trends Ecol Evol. 1996; 11:158–162. [PubMed: 21237791]

Wertheim JO, Chu DK, Peiris JS, Kosakovsky Pond SL, Poon LL. A case for the ancient origin of coronaviruses. J Virol. 2013; 87:7039–7045. [PubMed: 23596293]

Wertheim JO, Kosakovsky Pond SL. Purifying selection can obscure the ancient age of viral lineages. Mol Biol Evol. 2011; 28:3355–3365. [PubMed: 21705379]

Team W. H. O. E. R. Ebola Virus Disease in West Africa - The First 9 Months of the Epidemic and Forward Projections. N Engl J Med. 2014



	# Total	# Sequenced	% ORF Coverage	x Coverage	% bp >Q32	# S seg.	# L seg.
Batch 1	95	83	100% [99.7%-100%]	172 [14-17,227]	98.4%	95	95
Batch 2	116	113	99.9% [98.0%-100%]	200 [6-23,924]	97.6%	116	84
Total	211	196	99.9% [98.0%-100%]	181 [6-23,924]	97.9%	211	179
Matched	179	167	100% [98.0%-100%]	248 [14-23,924]	98.3%	179	179

Figure 1. Lassa fever is a viral hemorrhagic fever endemic in West Africa where Ebola virus disease broke out in 2014

(A) Overview of the LF endemic zone. Study sites are marked. The LF risk zone was defined according to Fichet-Calvet *et al.* (Fichet-Calvet and Rogers, 2009). (B) Schematic of LASV virions. (C) Summary of LASV sequence data (% ORF Coverage = average coverage of open reading frames; x Coverage = median base pair (bp) coverage; % bp > Q32 = fraction of bp with a phred-score > 32. (D) Plot of the combined normalized (to the sample average) genome coverages (Matched dataset, n = 167). See also Figure S1 and Table S1.

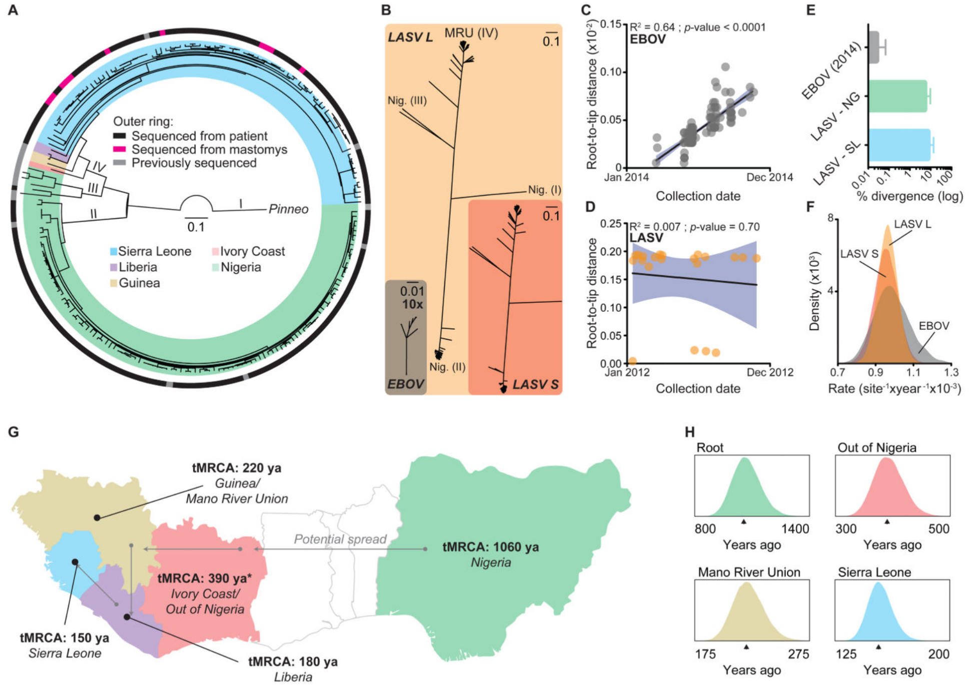


Figure 2. LASV is more diverse than EBOV and has ancient origins in Nigeria

(A) Phylogenetic tree of LASV S segments (n = 211) (outer ring: gray = previously sequenced; orange = sequenced from *M. natalensis*; scale bar = nucleotide substitutions/site; I – IV = lineages as defined by Bowen *et al*. (Bowen et al., 2000)). (B) Scaled trees of LASV L and S segments, as well as EBOV. Trees are shown with the same scale of genetic distance (0.1 nucleotide substitutions/site), except for EBOV, which was magnified 10× (0.01 nucleotide substitutions/site). LASV lineages are shown (Nig. = Nigeria; MRU = Mano River Union). (C, D) Root-to-tip distance versus collection date for (C) EBOV from the West African EVD epidemic (2014; n = 131), or (D) LASV from Sierra Leone (2012; n = 21). Confidence intervals (95%) for linear regression fits are shown in blue. (E) The % pairwise differences (log scale) in EBOV lineages from the 2014 EVD epidemic (March-October, 2014; n = 116) and LASV lineages from Sierra Leone (SL; 2009-2013; n = 60) and Nigeria (NG; 2009-2012; n = 83). The % divergence was calculated within the countries for each year separately and pooled. Error-bars represent the standard deviation. (F-H) Bayesian coalescent analysis of LASV samples (Matched dataset, n = 179). (F) Substitution rates. (G) LASV L segment tMRCA for each country (median values; ya = years ago). Gray arrows depict the likely spread of LASV. * = This tMRCA was dependent on only one sequence (AV) from outside Nigeria and MRU. (H) Probability distributions for the estimated tMRCAs with median marked. See also Figures S1F-H, S2-S4, Table S2, and Files S1-S3.

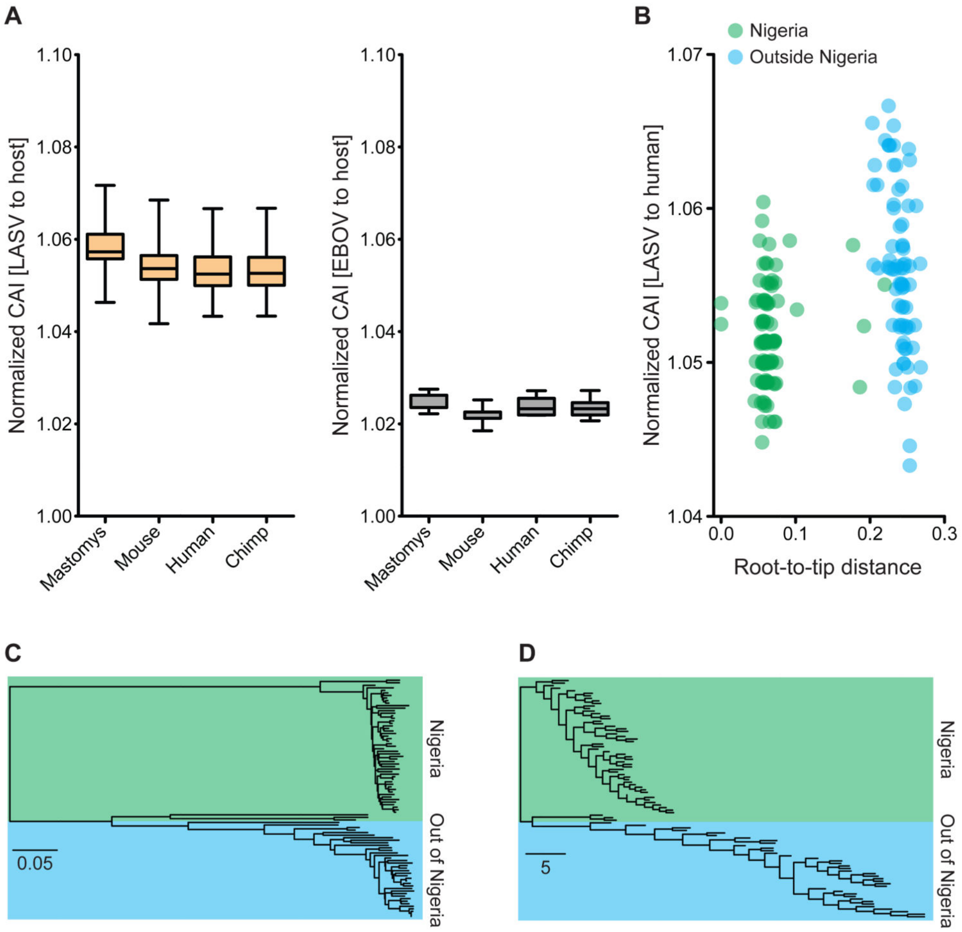


Figure 3. Increased codon adaptation of non-Nigerian LASV strains

(A) Codon adaptation index (CAI) of individual LASV (orange) and EBOV (gray) sequences to four mammalian hosts, normalized by GC and amino acid content. (B) Normalized CAI (to human) of LASV sequences plotted against their distance (aa substitutions/site) to the root of the tree. (C) Phylogeny of the LASV L genes (scale bar = substitutions/site). (D) A phenogram depicting the phylogeny from C with branch lengths representing CAI (scale bar = converted Z-score). (C, D) Trees were rooted on Pinneo (not shown; Batch 1 dataset). See also Figure S5A-H.

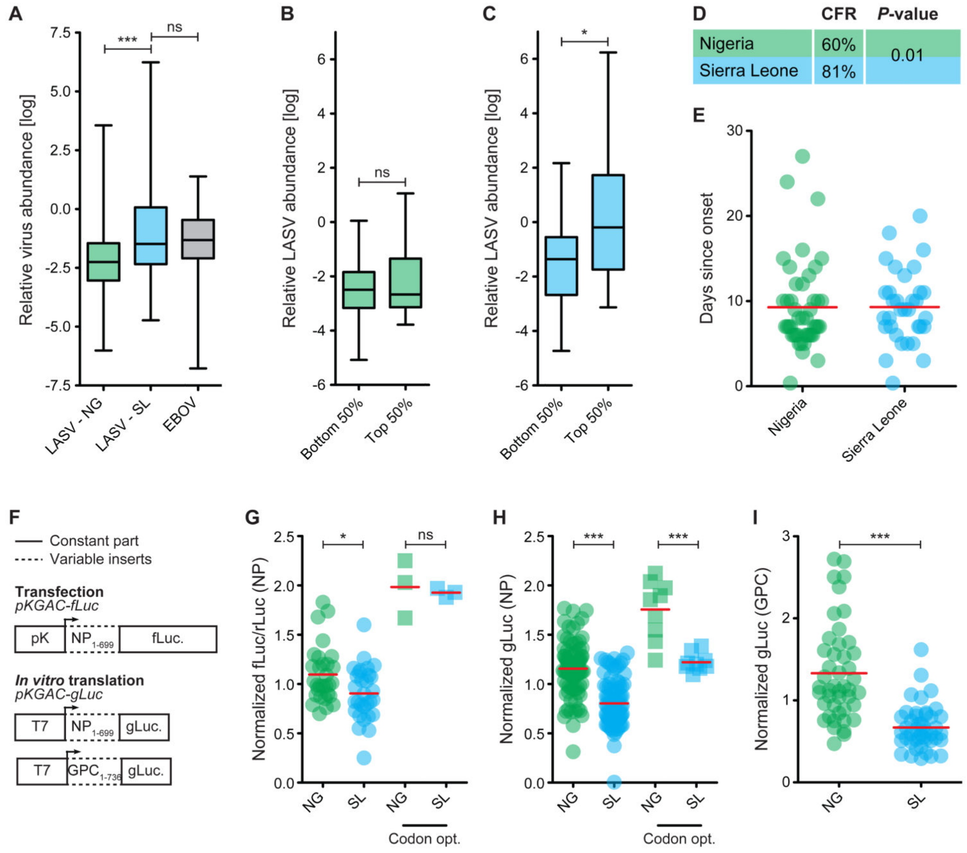


Figure 4. Difference in viral output between Nigerian and Sierra Leonean LASV strains

(A) Relative abundance of LASV and EBOV genome copies (log ratio of LASV or EBOV copies/μl to 18S rRNA copies/μl; *** = *P*-value < 0.001, Mann-Whitney test). (B, C) Relative abundance of LASV genome copies when partitioned into sequences in the top or bottom half of CAI scores. (B) Samples from Nigeria. (C) Samples from Sierra Leone (* = *P*-value < 0.05, Mann-Whitney test). (D) Case-fatality rates calculated for patients from Sierra Leone (n = 67) and Nigeria (n = 40). *P*-values from Fisher's exact test. (E) Patient-reported days from the onset of symptoms until admission to the hospital. Mean values are displayed with red bars. (F) DNA plasmids encoding the first 699 nucleotides of LASV NP or the first 736 nucleotides of LASV GPC. (G) NP-reporter expression was measured in HEK293 cells by the ratio of fLuc/rLuc 20 hours post transfection. (H, I) *In vitro* transcription of (H) NP- or (I) GPC-reporter translation measured by gLuc luminescence after 21 hours. (G-I) All values were normalized to the average of each biological replicate (n = 3). * = *P*-value < 0.05, *** = *P*-value < 0.0001, Mann-Whitney test; NG = Nigeria, SL = Sierra Leone. See also Table S2.

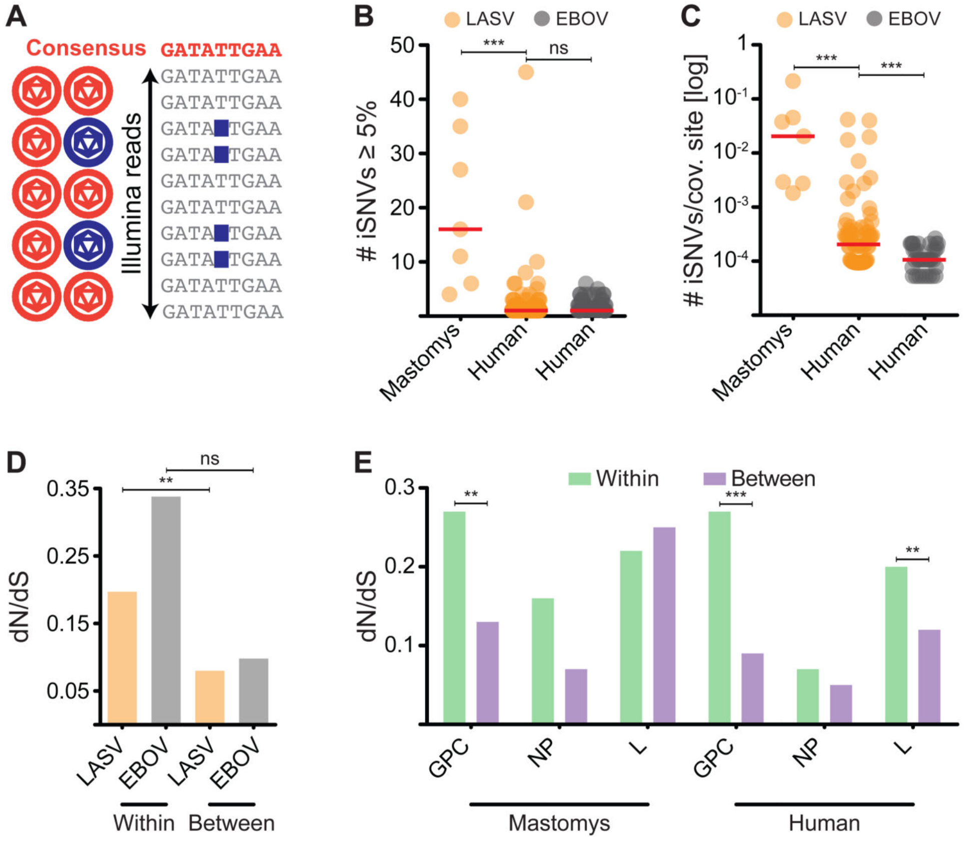


Figure 5. Genetic diversity and selective pressures within and between hosts

(A) iSNVs can be detected in Illumina reads. (B) Number of iSNVs at 5% MAF or higher, called in LASV (orange) and EBOV (gray). (C) Normalized number of iSNVs per covered site (80% calling power; 5% MAF). In both (B) and (C) each circle represents one LASV or EBOV sample; red bars denote the median; *** = *P*-value < 0.001, Mann-Whitney test; ns = not significant. Only samples with > 50× coverage were included. (D) dN/dS ratios for LASV and EBOV based on iSNVs or fixed differences in consensus sequences between hosts. (E) dN/dS ratios for each LASV gene. (D-E) *** = *P*-value < 0.001, ** = *P*-value < 0.01, Permutation test; ns = not significant. The very short Z gene was excluded. See also Figures S5J-L, S6, and S7, Table S2, and File S1.

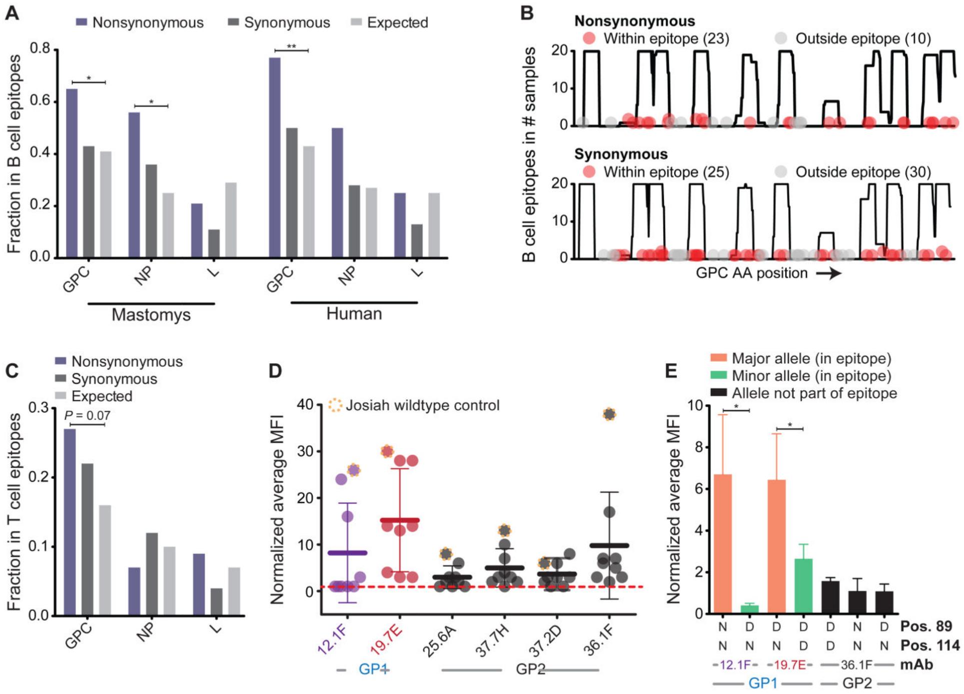
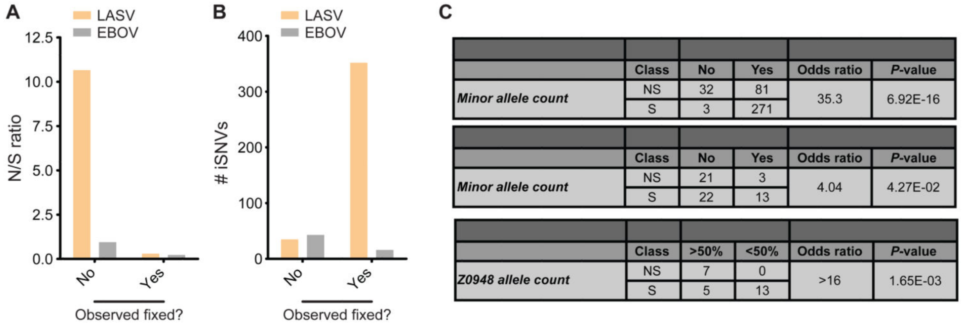


Figure 6. Nonsynonymous iSNVs are over-represented within predicted B cell epitopes in LASV GPC

(A) Fraction of iSNVs within predicted B cell epitopes. The observed fraction is compared to the expected fraction (** = *P*-value < 0.01, * = *P*-value < 0.05, Binomial test). (B) Overlap between GPC epitopes and iSNVs. Epitopes were predicted separately in each sample (y-axis) and overlaid with iSNVs from that sample. (C) Fraction of iSNVs falling within predicted T cell epitopes (*P*-value = Binomial test). (D, E) Binding of monoclonal antibodies (mAbs) to iSNV mutants in predicted B cell epitopes was tested in HEK293 cells. (D) Each circle correspond to the normalized average mean fluorescence intensity (MFI) measured by flow cytometry of each LASV GPC construct carrying either wildtype or iSNV mutations (Extended Experimental Procedures). Each tested mAb is shown on the x-axis. The MFI was normalized to the MFI of the empty vector control for each experiment. (E) Binding to the GP1-specific mAbs 12.1F and 19.7E using constructs carrying either the major or minor population-wide allele at positions 89 and 114. For comparison, binding to mAb 36.1F, which requires GP2, is also shown. All MFI values were normalized to the MFI of binding to the GP2-specific mAb 37.2D. (D, E) Error-bars show the standard-deviation from four independent experiments; * = *P*-value < 0.05, Mann-Whitney test. See also Figure S7I, Table S2, and File S1.



	Class	No	Yes	Odds ratio	*P*-value
Minor allele count	NS	32	81	35.3	6.92E-16
	S	3	271		

	Class	No	Yes	Odds ratio	*P*-value
Minor allele count	NS	21	3	4.04	4.27E-02
	S	22	13		

	Class	>50%	<50%	Odds ratio	*P*-value
Z0948 allele count	NS	7	0	>16	1.65E-03
	S	5	13		

Figure 7. Biased fixation of nonsynonymous iSNVs

(A) iSNVs that are never observed as fixed differences between consensus sequences have a higher N/S ratio in both LASV and EBOV. (B) LASV iSNVs are more commonly seen as fixed differences than EBOV iSNVs. The data displayed in (A) and (B) are tabulated in the top two panels of (C). (C) Biased fixation of iSNVs at the population-wide level (‘All iSNVs’) and in a pair of *M. natalensis* (‘Z0947 iSNVs’). At the population level (top and middle tables), the ‘Fixed?’ column indicates whether or not the minor iSNV allele is observed in any other LASV (top) or EBOV (middle) consensus sequences. The ‘DAF’ columns indicates the derived allele frequency in Z0947, with derived/ancestral allele states inferred from Z0948. [1]Fixation criterion: the minor iSNV is fixed (100%) in one or more other consensus sequences. [2]The DAF is defined as the frequency in Z0947 of the allele not fixed in the Z0948 consensus sequence. See also Figure S7H.